DOCUMENT RESUME

ED 425 645 FL 025 562

AUTHOR Hargett, Gary R.

TITLE Assessment in ESL & Bilingual Education: A Hot Topics Paper.

INSTITUTION Northwest Regional Educational Lab., Portland, OR. Comprehensive Center, Region X.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE 1998-08-00
NOTE 37p.
CONTRACT S283A50041
PUB TYPE Reports - Descriptive (141)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Standards; *Bilingual Education; Classroom Observation Techniques; Cloze Procedure; Educational Trends; *English (Second Language); Evaluation Criteria; Evaluation Methods; Intelligence Tests; Language Proficiency; *Language Tests; *Limited English Speaking; Oral Language; Reading Tests; Second Language Programs; Special Education; Speech Skills; Student Evaluation; Test Construction; Test Format; Test Items; *Testing; Trend Analysis; Vocabulary Development; Writing Exercises; Writing Tests
IDENTIFIERS *Placement Tests

ABSTRACT

The purposes and methods of testing in bilingual and English-as-a-Second-Language (ESL) education are discussed. Different instruments, including specific published tests, are listed and described briefly. They include language proficiency assessments, achievement tests, and assessments in special education. Introductory sections address topics surrounding the testing itself, including the need to understand the purposes of the testing and of a specific test, the information needed about a student, specific uses of that information, when it is appropriate to test the student, and defining second language proficiency. (Contains 20 references.) (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 425 645

Assessment in ESL & Bilingual Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A Hot Topics Paper

Northwest Regional Educational Laboratory's
Comprehensive Center, Region X

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Jerry D Kirkpatrick
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

BEST COPY AVAILABLE

© 1998 NWREL, Portland, Oregon

Permission to reproduce in whole or in part is granted with the stipulation that the Northwest Regional Educational Laboratory's Comprehensive Center, Region X, be acknowledged as the source on all copies.

The contents of this publication were developed under Cooperative Agreement Number S283A50041 with the U.S. Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and endorsement of the contents by the federal government should not be assumed.

ASSESSMENT IN ESL & BILINGUAL EDUCATION

A Hot Topics Paper

Gary R. Hargett, Ph.D.

NWREL's Comprehensive Center, Region X

August 1998

Northwest Regional Educational Laboratory
101 S.W. Main Street, Suite 500
Portland, Oregon 97204

Table of Contents

Introduction 1

Asking the Right Testing Questions 1

What Information is Needed About the Student? 2

Appropriate Courses of Action 4

When to Assess 5

Assessing Language Proficiency 6

Defining Second-Language Proficiency 6

Five Suggestions. 8

Instruments for Assessing Language Proficiency 9

Achievement Testing 23

Special Education 26

Conclusion 27

Bibliography 29

Introduction

Assessment plays a key role in every aspect of programs for limited-English-proficient (LEP) students. It figures in identifying the students who need those programs, placing them in the right levels of service, monitoring their progress, improving the programs that serve them, and deciding when a special program is no longer needed. This paper offers a nontechnical discussion of why, when, and how to assess LEP students. It does not pretend to be exhaustive, but it provides a foundation in the issues program planners should consider in selecting their assessment approaches.

The paper is structured around the purposes of testing in bilingual education and English as a second language. Different instruments, including specific published tests, are listed and briefly described for language proficiency assessment, achievement testing, and assessment for special education. However, readers should take time to read the sections of the paper on "Asking the Right Testing Questions," "When to Assess," and "Defining Second-Language Proficiency" in order to put those instruments in perspective. In fact, readers are encouraged to take the time to read the entire paper because it embeds various theoretical, instructional, and data-management issues within the context of different instruments and assessment strategies.

Two terms used throughout the paper have sparked some controversy: "limited English proficient (LEP)" and "alternative program of instruction." Many people object to the term limited English proficient because it defines students in terms of their limitations, what they cannot do as opposed to what they can do. The term "English-language learner" has begun to replace it, and occasionally, it is used in this paper. However, LEP is used here because it is commonly recognized and is used in the Improving America's Schools Act and other federal and state official documents. An "alternative program of instruction" for LEP students does not imply alternative education in the same sense that it is often used for at-risk students. The term is used by the U.S. Office for Civil Rights and in this paper to refer to strategies that overcome barriers to equitable education based on language differences. The "alternative program" refers to instructional strategies, not to curriculum. LEP students are entitled to access to the same curriculum and other programs and services as all other students.

Asking the Right Testing Questions

"Asking the right questions" means being clear about what purposes a test will serve before it is selected and used. Two "what" questions should guide the selection process:

- What information about a student is needed?
- What will that information be used for?

These two "what" questions must be answered before the more common "how" questions teachers usually ask regarding how to assess English proficiency, how to assess native-language literacy, how to assess academic achievement, etc.

What Information Is Needed About a Student?

Tests provide information about a student that is not otherwise available from records or simple observation. They also document that information for an audience—such as the public, parents, or policymakers—in a way that makes sense to the audience. In either case, a test provides certain kinds of information, and we cannot wisely select our assessment strategies for either purpose unless we know what that information is. When we want to assess students who are limited in their English proficiency, examples of the kinds of needed information may include:

- Can the student participate in the oral language of a mainstream classroom?
- Can the student read and write English at levels similar to his or her mainstream grade mates?
- Does the student need an ESL or bilingual program?
- If the student needs an ESL or bilingual program, what should his or her placement be?
- Does the student read and write the native language at grade level?
- Are the student's academic skills near grade level in the native language?
- What specific aspects of English grammar or vocabulary does the student lack?
- Is the student progressing in oral or written English?

The trick becomes matching the right assessment strategy to the school's purposes, that is, getting the answers to the school's questions about the student. The right strategy might be a published test, application of a scoring rubric, or simply the teacher's informed observation of a student's performance. In any case, the strategy must meet this one criterion: The assessment tasks must represent the kind of information the school wants to know about the student. Each of these kinds of information places certain requirements on the test. Figure 1 illustrates this point by listing requirements a test should meet to answer the questions listed above.

Figure 1

Requirements a Test Should Meet

Questions	Requirements
Can the student participate in the oral language of a mainstream classroom?	The tasks must simulate the oral language of a mainstream classroom.
Can the student read and write English at levels similar to his or her mainstream grade mates?	Ask the student to read or write something at that level.
Does the student need an ESL or bilingual program? If the student needs an ESL or bilingual program, what should his or her placement be?	Give tasks that represent a range of difficulty, from grade-level performance to little or no English proficiency.
Does the student read and write the native language at grade level?	Staff need to know what that language's grade-level expectations are; tasks should require the student to read and write something at that level.
Are the student's academic skills near grade level in the native language?	Give tasks that represent that language's grade-level standards in the academic areas of interest.
What specific aspects of English grammar or vocabulary does the student lack?	Tasks should pinpoint specific grammatical structures or vocabulary.
Is the student progressing in oral or written English?	Tasks should cover a range of difficulty, spanning at least the student's initial ability level and the level he or she is expected to reach after instruction.

- If the school wants to know whether the student can participate in the oral language of a mainstream classroom, the tasks must simulate the oral language of a mainstream classroom
- To find out whether the student can read and write English at levels similar to his or her mainstream grade mates, ask the student to read or write something at that level
- To find out whether the student needs an ESL or bilingual program, and what his or her placement should be, give tasks that represent a range of difficulty, from grade-level performance to little or no English proficiency

- If the school wants to find out whether the student reads and writes the native language at grade level, staff need to know what that language's grade-level expectations are; tasks should require the student to read and write something at that level

- To find out whether the student's academic skills are near grade level in the native language, give tasks that represent that language's grade-level standards in the academic areas of interest

- To find out what specific aspects of English grammar or vocabulary the student lacks, tasks should pinpoint specific grammatical structures or vocabulary

- If the school wants to know whether the student is progressing in oral or written English, tasks should cover a range of difficulty, spanning at least the student's initial ability level and the level he or she is expected to reach after instruction

Assessments that tell whether a student is limited English proficient, or performing similarly to age or grade mates in English or another language, are explicitly or implicitly standard-referenced. In other words, they set criteria for success according to the assessment's purpose (for example, the child is proficient in English, the child reads and writes his or her native language at grade level, etc.). Assessments that pinpoint specific weaknesses are not standard-referenced but offer diagnostic capabilities. Some of the previous questions imply standards, and some imply the need for diagnostic information. Test selectors must know which kind of information they want and select the right instrument accordingly. Examples of both kinds of instruments are given later in this paper.

Appropriate Courses of Action

Assessment information is not valuable if it does not lead to a course of action for the student's (and teacher's) benefit. For ESL and bilingual education, courses of action include:

- Placing the student in a program that develops his or her English while providing academic instruction

- Placing the student at the correct level of an ESL curriculum

- Providing the student with instructional services that match his or her academic preparation in the native language

- Reassigning the student into a different component or level of the alternative program according to new assessment results

- Improving the ESL or other instructional services if students do not show progress toward English proficiency or academic achievement
- Designing instruction to address specifically identified weaknesses

All of these applications of assessment information presuppose an alternative program of instruction for a school's LEP students. All assessment decisions should be made in the context of such a program that systematically identifies potential LEP students, assesses their English proficiency, places them in appropriate levels of service, monitors their progress in English and academic skills, and exits them when their English proficiency warrants. Readers who do not have such a program in place should refer to the Northwest Comprehensive Assistance Center's Bilingual/ESL Education Project, which provides guidance on legal requirements for serving LEP students and describes various program options. It also details legal requirements regarding assessment of LEP students.

When to Assess

The decisions on when to assess are guided partly by legal requirements and partly by good educational practice. The various assessment strategies described in this paper should be selected according their appropriateness for different purposes within an alternative program of instruction for LEP students.

Initial assessment should occur when potential LEP students are identified as they first register for school. They should be assessed for English-language proficiency in listening, speaking, reading, and writing to determine whether lack of English proficiency is likely to affect their opportunity to learn, indicating a need for an alternative program of instruction. For initial assessment, select an English proficiency test or assessment strategy that represents a range of difficulty, from age- or grade-level performance to little or no English proficiency. If using a published test that states guidelines for determining limited English proficiency or English fluency, look for evidence that those guidelines were established through field testing of the instrument.

Good teaching involves ongoing assessment to make sure students are progressing, whether in English proficiency, English or native-language literacy, or academic subjects. Ongoing assessment provides formative information for teachers to use in designing instruction to address specifically identified weaknesses. When standards and expectations are clearly communicated to students, it also allows students to become partners in assessing their own progress. Ongoing assessment can use less formal instruments than those used initially in identifying LEP students.

LEP students served in alternative instructional programs should be assessed annually for progress in English proficiency. Instruments and procedures used should, to the extent possible, be standard across the program; this allows the aggregation of data across grades and buildings to communicate to the public, parents, or policymakers. It also provides program leaders with patterns of success or failure among program components so they can improve instructional services. Instruments and procedures used in the annual assessment do not have to be the same ones used for initial LEP identification. They should, however, have similar qualities—i.e., they should represent a range of difficulty, from age- or grade-level performance to little or no English proficiency—so the program can show growth from initial assessment and track the student toward full proficiency and eligibility for program exit.

Assessing Language Proficiency

How to assess English proficiency is the main question facing ESL and bilingual educators, once they have addressed the "what" questions. They are legally obligated to assess the English proficiency of students with a primary or home language other than English to determine whether they are LEP and to provide appropriate services if they are so identified. Schools also are obligated to systematically monitor the students' progress in English. At the classroom level, teachers want to know students' "levels" of proficiency in order to gear instruction appropriately; they want to know what specific aspects of English different students are finding problematic in order to tailor instruction; and they want evidence that, in fact, their students are making progress.

Several instruments and strategies are available to meet these various purposes. Some are commercially published, some are available in the public domain, and some can be constructed by teachers. They rely on different kinds of logic for scoring, and often they represent different facets of English proficiency. These instruments and strategies are most effective when administered and interpreted by teachers or other staff people who have insight into what we are measuring when we measure proficiency in English as a second language. Therefore, it is important to take a few minutes to develop an understanding of what language proficiency means.

Defining Second-Language Proficiency

Everyone agrees that language is used for communicating thoughts, ideas, or information. Communication requires at least two people, and it requires that those two people have a common system that regularly represents the thoughts, ideas, or information they wish to share through language; one speaks or writes, the other listens or reads.

Anyone who has studied or even encountered a second language immediately recognizes that the second language differs from his or her own in that it uses a different vocabulary to represent individual thoughts, ideas, or information. A person who has tried to learn a second language solely by learning word lists soon comes face to face with the reality that simply stringing words together does not necessarily communicate. Sometimes words in a second language appear in a different order from the native language, and many other rules govern the relationships among words.

For example, different languages attach very different inflections to the ends of verbs to indicate past time, while some do nothing at all to the verbs. Each language has a systematic grammar, shared among its users, that clarifies the relationships among the words in an utterance. Every natural human language gives physical form to its words through speech sounds (with a few notable exceptions, such as American Sign Language). Different languages use different sounds, and different languages may use the same sounds differently. The sounds of a language are also systematically governed by rules of phonology that are shared by its speakers. The most commonly encountered languages also represent their words in written forms, and written languages also follow elaborate rule systems.

To be proficient in a second language means to effectively communicate or understand thoughts or ideas through the language's grammatical system and its vocabulary, using its sounds or written symbols.

Any or all components of this definition may comprise the kind of information you wish to learn about students when assessing their language proficiency. In one way or another, language proficiency tests assess the student's communication ability or knowledge of the components of language. When deciding how to assess, educators must first decide which aspect of language proficiency they are interested in. Those aspects might include the ability to communicate orally and in writing similarly to non-LEP grade mates; to use correct or clear pronunciation; or to exhibit knowledge of specific grammatical structures. They must then select the appropriate assessment tool specifically designed to obtain that information.

In recent years, most ESL or bilingual educators have taken a stronger or weaker form of the position that language cannot be "taught." Rather, its acquisition can only be facilitated. What is important, then, is not discrete grammar or vocabulary but communicative competence. Individual educators may make up their own minds on this point. However, our definition of language proficiency shows that effective communication happens with command of a language's grammar and vocabulary, and spoken or written representations of them, whether they can be explicitly taught or not. It is also important to remember that current educational reform movements set standards for writing, and correct use of English conventions figures prominently in those standards. English-language

learners may produce well-organized and interesting writing, but they will likely be penalized for the kinds of errors that stem from incomplete command of English's linguistic rules.

Five Suggestions

Before considering the various instruments and strategies available for assessing language proficiency, five suggestions are offered that may affect how they are used.

First, many of these instruments—even ones that have quite specific instructions for the examiner—rely to an extent on the examiner's personal judgment in scoring an item right or wrong or assigning a rating. Remember that a fair test is not necessarily one on which the student can perform well and get a high score; a fair test is one that accurately assesses a student, providing a score that reasonably represents the student's true ability. In individually administered tests, there's often a temptation to prompt the student, to give credit for being "in the ballpark," or to say, based on other interactions with the student, "I know he really knows that." This attempt to be "fair" can in fact penalize the student by failing to identify him or her for needed services. Follow established testing procedures as rigorously as possible. Otherwise, the test may not give a true reading, and will mean different things for different students, undermining the purpose for giving the test to begin with.

Second, several of the instruments or procedures have other-language forms or can easily be adapted for other languages besides English. It is useful to assess students' abilities in their native language, especially when the alternative program has the capability to provide native-language instruction and development. The fact that a student with a home language other than English is limited in English does not mean that student is proficient in the home language. The student may profit from additional native-language development. In some cases, it may be that the student's English, though limited, is stronger than the home language. Two students, testing at the same level in English and with the same native language, may be candidates for different kinds of program services depending on the strength of their native language.

Third, do not rely on one instrument to meet all needs. Select a variety of instruments or strategies for the most complete picture of a given student or a group of students. Instruments that work well for initial identification and program exit may not be useful in helping a teacher plan for a given class of students or for individualizing instruction. Similarly, instruments that provide rich individual information may not lend themselves to aggregation for program evaluation or to communicating progress to a public audience. And sometimes an instrument gives information that doesn't match other sources of information. If a student

tests as LEP on an English proficiency test but reads and writes English as effectively as monolingual English-speaking grade mates, the latter information might override the test score.

Fourth, ESL and bilingual program staff should enlist the support of the district's testing specialists. Initial identification and annual progress assessment especially should be integrated into the district's broader assessment program. It is a mistake to leave ESL and bilingual testing solely to the personnel assigned to that program. Appropriate services for LEP students, including identification and monitoring, are the district's legal responsibility. Integration helps institutionalize that responsibility. Also, ESL and bilingual program staff may not have the technical expertise needed to interpret and report testing data. For example, some tests provide different kinds of scores—such as level scores, scale scores, and percentiles—which are analyzed and reported in unique ways. The district's testing specialists should be involved in the analysis, interpretation, and reporting of programwide data.

Fifth, make sure all educational personnel who serve LEP students understand the implications of the assessment results and what limited English proficiency means. It does not mean the student lacks innate ability or cannot learn. In many cases, it does not mean the student is not at grade level; many LEP students have records of academic success, but they cannot demonstrate their knowledge in English. The student's identification as LEP simply means the student needs help in acquiring English proficiency and should be taught in ways that account for his or her linguistic differences.

Instruments for Assessing Language Proficiency

This discussion of language proficiency assessment instruments concentrates on English proficiency assessment because that is the one constant in all ESL and bilingual education programs. Some of the instruments have non-English forms, and some can easily be adapted to other languages. Implications and applications for other languages are discussed in the context of each instrument. Many more instruments exist than are listed here. The tests selected for inclusion in this paper illustrate the matching of tests to various purposes, represent different approaches to language assessment, and are commonly used among ESL or bilingual programs in the Northwest.

Unfortunately, instruments for assessing language proficiency do not lend themselves to any logical taxonomy. It would be very helpful to say, "These are the instruments to assess oral proficiency; here are pragmatic, integrative tests of communicative competence; this list contains discrete-point tests that isolate problematic grammar; this test is appropriate for elementary school program

placement, that one for secondary schools." In real life, the instruments spill over across categories and imperfectly fit others, so this discussion will consider each instrument in turn and discuss what it assesses, how it is scored, and how it can be used to answer questions about students' language abilities.

This discussion is not intended to prepare the reader to administer the instruments, nor to evaluate their technical quality, nor to corroborate or dispute their claims. For published instruments, technical or examiner manuals are available to provide technical data and detailed instructions on administration, scoring, and interpretation. Readers who are interested in a specific instrument should request additional information from the publisher and determine for themselves whether it suits their programs.

In some cases, examples are given of how a test's data can be reported. Those examples are not necessarily unique to that test. For example, if an example is given of the reporting of NCE data, that example could apply to any test that yields that kind of score.

Language Assessment Scales, Oral (LAS-O)
Published by CTB\MacMillan\McGraw-Hill

The LAS-O is published in three different forms: the Pre-LAS, intended for young children approximately four to six years of age; the LAS-O I, for elementary grade levels; and the LAS-O II, for secondary levels. (A LAS-A, for adult second-language learners, is also available.) The LAS-O is individually administered. It presents a variety of oral language tasks, including supplying correct vocabulary for pictures of objects; responding to comprehension questions or tasks; and producing oral language in response to picture stimuli. Four categories of tasks are discretely scored on the basis of right or wrong answers. A fifth category, accounting for half the score, utilizes a rating scale.

The LAS-O is scored to classify students into five different proficiency levels subsumed within the broader designations of non-English speaking, limited English speaking, and fluent. It therefore can meet a number of purposes. It can be used as part of the initial identification of LEP students needing an alternative program of services. Depending on how the alternative program is structured, it can place students into different instructional groupings. It can be used to track annual progress in oral English proficiency. And it can form part of the criteria for redesignating students as no longer limited in oral English proficiency and possibly eligible for program exit.

The LAS-O provides two principal kinds of scores: proficiency levels (1-5) and converted scores (1-100). The proficiency levels are useful for categorizing students as eligible for the ESL or bilingual program, and for producing data that

depict the characteristics of a school's LEP population. For example, they can be used to construct a table such as that in Figure 2, drawn from a small program serving 50 students in kindergarten through grade three. The figure shows the distribution of the students by grade and LAS proficiency levels. The reader can tell at a glance that initially, low-proficient students were more concentrated at lower grade levels. A visual comparison of the pretest and posttest data shows there was general progress because more students at all grade levels scored at higher proficiency levels on the posttest.

Figure 2

Sample Display of LAS Level Data
English LAS Pretest Levels

Grade Level	K	1	2	3
1	10	7	3	2
2	4	2	1	
3	7	2	1	7
4	4	2		
5	5	2		

English LAS Posttest Levels

Grade Level	K	1	2	3
1	6	3		2
2	2	2		
3	5	2	3	6
4	7			
5	1	6	2	2

However, proficiency-level scores have limitations. They can understate or overstate student progress, particularly for individual students. A level score spans a range of converted scores. For example, converted scores of 55 and 63 both fall within proficiency level 2. But based on level scores, a student who advanced from 55 to 63 would show no growth. Similarly, a student might gain only one or two points, essentially making no progress, yet could show advancement from one proficiency level to another. It is a good idea to collect, record, analyze, and report both kinds of scores.

The LAS-O is also available in Spanish. It is useful to assess students' abilities in their native language, especially when the alternative program has the capability to provide native language instruction and development. The fact that a student with a home language other than English is limited in English does not mean that student is proficient in the home language. The student may profit from additional native-language development, and in some cases, it may be that the student's English, though limited, is stronger than the home language.

Language Assessment Scales, Reading and Writing (LAS R/W)
Published by CTB\MacMillan\McGraw-Hill

The LAS R/W is published in three different levels to accommodate grades two through high school. Each of the three levels contains a variety of reading and writing tasks, some scored on a right-or-wrong basis, some judged according to a rating scale. Scores on the different sections are combined into standard scores, 0-100, which in turn can be interpreted according to three reading/writing competency levels, representing non-reader/writer; limited reader/writer; and competent reader/writer. All versions of the LAS R/W can be group administered. It is available in both English and Spanish.

Because the LAS R/W provides scores that indicate whether a student is performing at a competent level in reading and writing, and it spans a range from non-reader/writer to competent, it can be used as part of the initial identification process, to track annual growth, and as one of the criteria for exit from an alternative program of instruction.

Woodcock-Muñoz Language Survey
Published by Riverside Publishing Co. (Houghton Mifflin)

The Woodcock-Muñoz Language Survey is intended to provide information on a student's cognitive and academic language proficiency, that is, the extent to which the student commands the kind of language typically required in school. It is individually administered. It has both English and Spanish forms, both consisting of four subtests, two of which comprise a score for oral language ability, two a reading and writing ability score. Together, all four comprise a broad language-ability score. All items are scored as right or wrong; there are no productive tasks—such as original writing or spoken discourse—scored on a rating scale. It is constructed to represent the language abilities expected at all ages or grade levels from preschool through college.

The Woodcock-Muñoz yields several different kinds of scores. It provides levels, called CALP (cognitive academic language proficiency) levels 1-5. The levels designate negligible English or Spanish; very limited English or Spanish; limited English or Spanish; fluent English or Spanish; and advanced English or Spanish. (Levels can be expressed at intermediate values; for example, a 3-4 is a student

between levels 3 and 4.) It also yields grade equivalents, age equivalents, percentiles, normal-curve equivalents, scale scores (called W scores), and standard scores.

When both the English and Spanish forms are used, they can provide a comparative language index, indicating which of the two languages a student has a better command of.

The Woodcock-Muñoz can be used as part of the initial identification of LEP students needing an alternative program of services. Depending on how the alternative program is structured, it can place students in different instructional groupings; it can be used to track annual progress in language proficiency; and it can form part of the criteria for redesignating students as no longer limited in English proficiency and eligible for program exit.

The variety of scores provided accommodates various interpretation and reporting needs. For example, if CALP levels are reported in both English and Spanish, and a school's program has bilingual capabilities, those scores can be used for placement in different components. Consider the hypothetical example of data from 71 students at the same, or perhaps adjacent, grade levels who took both the English and Spanish versions, shown in Figure 3 on the following page. Several students are clearly limited in English and much stronger in Spanish, making them logical candidates for native-language instruction that builds on their native-language abilities. Others are limited in both Spanish and English and could profit from further, intensive development in both. Some are limited in English, but their English is clearly stronger than their Spanish. Those students need an alternative program of instruction, but they may not be candidates for instruction in Spanish, particularly if resources for that are limited.

As stated previously, level scores have limitations. For example, they are difficult to use in tracking progress for groups of students. For that purpose, NCE scores or scale scores may be preferable. Figure 4 presents a hypothetical example of an elementary ESL program using Woodcock-Muñoz Broad English NCE averages by grade level to represent gains made over one year's time, from 1996-1997.

This kind of display can be used to show that by and large, students advanced in their broad English abilities but remained quite low and still in need of an alternative program of instruction.

Figure 3

Sample Display of Woodcock-Muñoz CALP Level Data in English and Spanish

Broad English Ability by Broad Spanish Ability

	English									
	1	1-2	2	2-3	3	3-4	4	4-5	5	Total
Spanish										
1						1				1
1-2	4	3	3	4						14
2			5	4	7					16
2-3				2	6					8
3	6					2				8
3-4	5	7	1							13
4	1	2		4			1	2		10
4-5									1	1
5										
Total	16	12	9	14	13	3	1	2	1	71

Figure 4

Sample Display of Woodcock-Muñoz Broad English NCE Gains

Grade	1996 Average	1997 Average	Gain	N
1	11.8	24.9	13.1	36
2	18.6	23.0	4.4	25
3	20.3	28.6	8.3	23
4	18.4	32.9	14.5	12
5	17.1	29.6	12.5	16
6	23.8	25.1	1.3	11

IDEA Oral Language Proficiency Tests (IPT)
Published by Ballard & Tighe

The IPT includes normed oral language proficiency tests for students prekindergarten through high school, with three levels in both English and Spanish. The Pre-IPT is intended for ages three through five; the IPT I for kindergarten through grade six; and the IPT II for grades seven through 12. All three provide designations for non-English or Spanish speaking, limited English or Spanish speaking; and fluent English or Spanish speaking. They are individually administered.

The designations of non-English speaking, limited English speaking, and fluent English speaking allow the IPT oral tests to serve as part of the initial identification of LEP students needing an alternative program of services system and the criteria for redesignating students as no longer limited in English proficiency and eligible for program exit.

IDEA Reading and Writing Proficiency Tests (IPT)
Published by Ballard & Tighe

The IPT Reading and Writing Proficiency Tests are group administered. They are available in three levels, the IPT I for grades two and three; IPT II for grades four through six; and IPT III for grades seven through 12. They are published in both English and Spanish versions and yield diagnostic reading profiles, percentiles, and NCEs. They can be used as part of the initial identification and program exit process.

Secondary Level English Proficiency (SLEP) Test
Published by Educational Testing Service

The SLEP, as the name implies, is intended for secondary students, approximately grades seven through 12. It consists of a listening comprehension and a reading comprehension section, both of which consist exclusively of multiple-choice items. Therefore, students should have experience with this kind of test format.

SLEP scores appear in both scale score and percentile forms. Designations are not provided for limited or fluent English proficiency, but guidance is given on what students in different score ranges can be expected to do. Districts can use these guidelines to establish local criteria. The SLEP can be used as part of initial LEP designation, to assess annual progress, and for consideration for program exit.

Student Oral Language Observation Matrix (SOLOM)

The SOLOM is not a test per se. A test is a set of structured tasks given in a standard way. The SOLOM is a rating scale that teachers can use to assess their students' command of oral language on the basis of what they observe on a continual basis in a variety of situations—class discussions, playground interactions, encounters between classes. The teacher matches a student's language performance in five domains—listening comprehension, vocabulary, fluency, grammar, and pronunciation—to descriptions on a five-point scale for each (see Figure 5). The scores for individual domains can be considered, or they can be combined into a total score with a range of five through 25, where approximately 19 or 20 can be considered proficient. SOLOM scores represent whether a student

can participate in oral language tasks typically expected in the classroom at his or her grade level.

Because it describes a range of proficiency from nonproficient to fluent, the SOLOM can be used to track annual progress. This in turn can be used in program evaluation, and as some of the criteria for exit from alternative instructional programs. However, to be used for these purposes, it is important to ensure that all teachers who use it undergo reliability training so that scores are comparable across teachers. For this purpose, a training video has been produced by Montebello School District in California.

The SOLOM does not require a dedicated testing situation. To complete it, teachers simply need to know the criteria for the various ratings and observe their students' language practices with those criteria in mind. Therein lies the greatest value of the SOLOM and similar approaches: It fixes teachers' attention on language-development goals; it keeps them aware of how their students are progressing in relation to those goals; and it reminds them to set up oral-language-use situations that allow them to observe the student, as well as provide the students with language-development activities. While observing, teachers should be attuned to the specific features of a student's speech that influenced their rating. They can use this information as a basis of instruction. The SOLOM is sufficiently generic to be applicable to other languages besides English.

The SOLOM is not commercially published. It was originally developed by the San Jose Area Bilingual Consortium and has undergone revisions with leadership from the Bilingual Education Office of the California Department of Education. It is within the public domain and can be copied, modified, or adapted to meet local needs.

Writing Samples

Writing is both an academic subject and a component of second-language proficiency. It should figure prominently as part of native-language development in bilingual programs and as part of English as a second language instruction. Students should write often, and teachers should assess their students' writing often. The assessment can and should take two different forms. One is the application of scoring guides (or rubrics), such as the six-trait Direct Writing Assessment used in Oregon's state student assessment program, or those contained in the LAS R/W. Scoring guides should be used periodically to monitor students' progress toward standards of good writing. It is a mistake to think that concerns for standards-based writing should be deferred until students become proficient in English.

Figure 5

SOLOM Teacher Observation

Student Oral Language Observation Matrix

Student's name_____ Grade_____ Date_____

Administered by (signature)_____ Language observed_____

	1	2	3	4	5
A. Comprehension	Cannot be said to understand even simple conversation.	Has great difficulty following what is said. Can comprehend only social conversation spoken slowly and with frequent repetitions.	Understands most of what is said at slower-than-normal speed with repetitions.	Understands nearly everything at normal speech, although occasional repetition may be necessary.	Understands everyday conversation and normal classroom discussions.
B. Fluency	Speech so halting and fragmentary as to make conversation virtually impossible.	Usually hesitant: often forced into silence by language limitations.	Speech in everyday conversation and classroom discussion frequently disrupted by the student's search for the correct mannef of expression.	Speech in everyday conversation and classroom discussions generally fluent, with occasional lapses while the student searches for the correct manner of expression.	Speech in everyday conversation and classroom discussions fluent and effortless; approximating that of a native speaker.
C. Vocabulary	Vocabulary limitations so extreme as to make conversation virtually impossible.	Misuse of words and very limited vocabulary: comprehension quite difficult.	Student frequently uses wrong words: conversation somewhat limited because of inadequate vocabulary.	Student occasionally uses inappropriate terms and/or must rephrase ideas because of lexical inadequacies.	Use of vocabulary and idioms approximate that of a native speaker.
D. Pronunciation	Pronunciation problems so severe as to make speech virtually unintelligible.	Very hard to understand because of pronunciation problems. Must frequently repeat in order to make him/herself understood.	Pronunciation problems necessitate concentration on the part of the listener and occasionally lead to misunderstanding.	Always intelligible, although the listener is conscious of a definite accent and occasional inappropriate intonation patterns.	Pronunciation and intonation approximate that of a native speaker.
E. Grammar	Errors in grammar and word order so severe as to make speech virtually unintelligible.	Grammar and word order errors make comprehension difficult. Must often rephrase and/or restrict him/herself to basic patterns.	Makes frequent errors of grammar and word order that occasionally obscure meaning.	Occasionally makes grammatical and/or word order errors that do not obscure meaning.	Grammar and word order approximate that of a native speaker.

Based on your observation of the student, indicate with an "X" across the category which best describes the student's abilities.

- The SOLOM should only be administered by persons who themselves score at level "4" or above in all categories in the language being assessed.
- Students scoring at level "1" in all categories can be said to have no proficiency in the language.

The Oregon Department of Education has produced a Spanish version of the Direct Writing Assessment scoring guide. The guide may be applicable in other languages as well.

Figure 6 shows a "downward extension" of the six-trait Direct Writing Assessment, which was developed out of concern that many students in an ESL program actually progress in writing, but the standard six-point scale does not accommodate their very low levels of writing. On the standard scale, they score at level 1, which actually may encompass several levels of ability. By adding levels 0 and minus 1 (–1), ESL program personnel can demonstrate the real progress students make. Moreover, the Direct Writing Assessment is not used below third grade. Therefore, the downward extension may be useful for students who have not attained the equivalent of third-grade skills, whether in English or in their own language.

State-adopted or locally produced scoring guides can play a role in the initial identification of LEP students, annual monitoring of progress, and exit decisions if they evaluate writing in relation to grade-level standards and are scored and interpreted by adequately trained personnel.

Writing samples can also provide information on very specific aspects of a student's language development. Teachers who are knowledgeable in language acquisition and descriptive or applied linguistics can use a student's writing to obtain a wealth of information. When ESL students write, their errors should not merely be corrected. They should be seen as signposts in their language development, showing the teacher what the student's personal English looks like. Our definition of second-language proficiency used the word "system." A second-language learner takes information about the target language and tries to decipher and internalize its system. Sometimes the learner gets it wrong, but without corrective feedback, whether through modeling or explicit instruction, the erroneous system may remain in place, especially if the learner successfully uses it for communication. If the goal is for the learner to master English at native-speaker standards, those errors need to be addressed.

Writing samples provide an excellent vehicle to detect and analyze errors. A high school ESL student wrote, "My friend and I decided to went to the mall to bought some new clothes." Often, ESL students neglect to inflect for the past tense at all. This student went the other way and inflected every verb in sight. She was very aware that English requires that actions in the past be signaled with the past-tense inflection, but she may have overgeneralized that rule, thinking (not necessarily consciously) that English demands consistency in verb tense. It does, but not in the way she apparently hypothesized. Many ESL programs teach about verb tense, but they do not cover every contingency of use. Grammar is

much too complex and intricate for a typical course to anticipate situations such as this girl produced, but the writing sample gave her teacher information for an immediately relevant and useful ESL lesson. When students write often, and knowledgeable teachers evaluate their writing, powerful ESL instruction can ensue.

Figure 6

"Downward Extension" of the Direct Writing Assessment

Ideas and Content 0 The writing begins to show a reason for communicating. • Dictates/draws stories • Associates words and pictures • Lists seem to reflect meaning -1 The writing has no meaning, just indicates learning the orthography. Expresses ideas with pictures. • Scribbles • Uses random letters or usually uses pictures only • No real meaning in print	**Sentence Fluency** 0 The student is beginning to associate meaning in a written format. He or she: • Can fill in blanks in a sentence pattern • Can copy a sentence • Can copy a list • Has a sense of sentence format -1 The writer has no sense of sentence format. • Copied words do not show spaces.
Conventions 0 The writer is beginning to understand that writing has a format that makes it express meaning. • Words are spaced • Visual memory of words listed • Phonetic spelling • Punctuation placed somewhere -1 The writer is beginning to form or copy letters.	**Word Choice** 0 The writing conveys meaning. • Uses letters to convey meaning • Prints recognizable words • Associates words and pictures -1 The writer is learning that ideas can be expressed with pictures. • Subsymbolic • No written words
Organization 0 The writing begins to show a pattern of a story. • Dictates/draws stories • List format • Linking of thinking to words -1 The writer is learning the orthography of writing. • Copied words do not show spaces	**Voice** There are no traits in 0 or 1.

Adapted from Juneau First-Grade Portfolio by Judith Perry for Portland Public School's ESL/Bilingual Program

The student whose writing appears in Figure 7 likewise is struggling with English verbs. In describing the new holiday she would inaugurate, she says, "The day I will celebrated in May 30." She should have said, "I will celebrate the day on May 30" or "The day will be celebrated on May 30." It would take further probing to know whether the problem lies in verb tense or formation of the passive voice. (She also does not understand the distinct uses of "in" and "on.") This student's oral English is clearly ahead of her written English. She uses the word "recognize" but does not spell it correctly; perhaps she has not encountered it in print or has not attempted to write it before. She also uses invented spelling for "teenagers."

Figure 7

Topic: *Create a new holiday. Give it a name and tell what date it will be celebrated. Describe how you will celebrate and explain why you chose it.*

I create a new holiday called Students day. The day I will celebrated in May 30,. I will celebrate going with my frinds and had a good day. I chose it because I think the student has to reconose they job. The students who study try to be someone in this world, and not all the tennegers study. I think reconose them is a very import thing

If states begin evaluating writing according to high standards at third grade, first- and second-grade teachers need to promote writing so their students will be ready when the time comes. Student writing at primary grades also provides rich insights into students' understanding of language, whether their second or first. Figure 8 shows the work of a second-grader who expresses her ideas well enough but is clearly still figuring out the relationship between Spanish as she speaks it and Spanish as it is written.

Figure 8

nancy/ 10 feb 997 /

un dia muy feliz. Yo jui a la puga I bi a mi prma I des pues yo ju: a la tieba I mi papa me copro un elado I des pues me gebo a comer a la comda china I me jebo a

One's personal native-language system may differ from the commonly accepted or conventional version, and this girl's writing shows evidence of that. She apparently knows word boundaries but seems to think of *despues* as two different words, perhaps because she knows that *pues* is a word. Some of her errors are typical of early writers in both English and Spanish, for example the backward "d" in *tienda*. Other errors are common among Spanish speakers even at more advanced levels of literacy, such as the substitution of "b" for "v" in *vi* and *llevó* (which she spells as "gebo"). She does not hear the nasal "n" or "m" as distinct consonants when they occur before another consonant, as seen in her spelling of *tienda* and *compró*, common among English speakers also. She does not neglect the "n" and "m"; she just does not hear them as distinct consonants. She writes Spanish as she hears and speaks it, spelling (and probably pronouncing) *fui* as "jui." Her Spanish is systematic, but it doesn't entirely follow the same system as standard Spanish.

School helps children achieve educated, standard forms of language. Writing samples provide detailed insight into how children's language differs from the standard form, whether in the first or second language. They also help teachers pinpoint aspects of grammar, vocabulary, and even pronunciation that need to be addressed in instruction.

Cloze Technique

Cloze testing is a handy technique for students who can read at a discourse level. Teachers can use it to determine whether a reading passage is plausibly within students' instructional range or prohibitively difficult. It can also provide useful information about specific language needs in grammar or vocabulary. Teachers can construct a cloze test by taking a reading passage they think is within the students' instructional range, leaving the first sentence intact, and then beginning with the second sentence, deleting every fifth, sixth, or seventh word. The students are given the passage with deletions and told to fill the blanks with a word that fits. For example, take a passage such as the one shown in Figure 9, and then delete every sixth word as shown in Figure 10.

Figure 9

Sample Passage for Cloze

Last weekend we went to the mountains to play in the snow.

The mountain roads were very slippery, and it was hard to drive.

It had snowed a lot the night before, so the snow was very deep.

At the end of the day when we tried to leave, we couldn't. The car was stuck!

Figure 10

Sample Cloze Deletions

Last weekend we went to the mountains to play in the snow.

The mountain roads were very _____ , and it was hard to _____.

It had snowed a lot _____ night before, so the snow was very _____.

At the end of the _____ when we tried to leave, _____ couldn't.

The car was _____!

Note that all the blanks are the same length regardless of the length of the words they replaced, to give the students room to write their responses and to avoid giving hints about what word belongs.

After the students fill in the blanks, if the passage doesn't make much sense, chances are the original passage would not have made much sense to them either, and it is probably above their instructional level. If the newly reconstructed passages generally make sense but contain words that do not fit the context, those words provide ideas for subsequent instruction. For example, if a student supplies "big" for "deep," that student needs vocabulary enrichment, particularly in adjectives related to weather.

The unusual format of cloze, and various technical problems associated with its validity and interpretation, make it unsuitable for initial identification of LEP students or for annual progress assessment. However, it can be a useful diagnostic tool, and used occasionally, an interesting class activity.

Dictation

Dictation can be used for diagnostic assessment with students who can read at a discourse level. Teachers can use it to determine whether a reading passage is plausibly within students' instructional range or prohibitively difficult. It can also provide useful information about specific language needs in grammar or vocabulary.

Teachers can construct a dictation by taking a reading passage they think is within the students' instructional range, and reading it to the class once all the way through at a normal instructional rate. The students are instructed simply to listen. Then they are instructed to write what they hear. The teacher breaks

the passage into short phrases, with natural phrasing, pausing between phrases to give the students time to write. Finally, the teacher reads the whole passage through at a normal rate to let students double-check their work.

If the students were able to reproduce the passage recognizably, it is most likely within their instructional range. Specific errors such as missing tense markers or words spelled incorrectly but phonetically may indicate points where students should receive targeted instruction. To the extent that the students could not reproduce many words or whole phrases, the passage was beyond their comprehension.

Dictation has good diagnostic capabilities and can provide an interesting occasional classroom activity. However, it is difficult to score and does not fit well as part of the initial identification process or annual progress assessment.

Achievement Testing

Achievement tests measure the extent to which students have mastered the knowledge and skills generally expected at their age or grade levels. School districts have the obligation of helping their LEP students overcome language barriers for a fair chance at mastering the same knowledge and skills expected of all other students. Therefore, they must also provide evidence that their LEP students are achieving in basic skills and core academic subjects. LEP students may be proficient readers and writers, although not in English, and they may command challenging subjects such as math, science, and history. However, their language differences prevent their demonstrating their ability on the same achievement tests in English that are used for other students, requiring ESL and bilingual programs to seek alternative means of measuring achievement.

Several strategies can be followed to include LEP students in achievement testing, and policies should be in place to ensure they are accounted for. Readers are encouraged to familiarize themselves with their own state or district policies on this issue. For example, Oregon has undertaken the development of non-English versions of its student assessments.

A detailed enumeration of policy and practice options for LEP achievement testing lies outside the scope of this paper, but options include extra testing time, giving directions in the students' native language, local development of native-language versions of tests aligned with content and performance standards, and adoption of achievement tests especially designed for LEP students, some of which are described on the following pages.

A word of caution is in order. The notion has arisen that performance assessments are somehow more valid for LEP students than more traditional paper-and-pencil or multiple-choice tests. Performance assessments provide valuable information, but they present many technical problems in construction. A performance assessment is not automatically more reliable or valid for a given group of students. Most performance assessments require students to use language extensively, implying they should be done in the native language if students are LEP. If a performance assessment were designed with a low language load—for example, building a model or producing an art project—it would be incumbent on the developers to explain the criteria used to claim the product represented a standard of achievement.

It goes almost without saying that the best way to find out if someone knows something is to ask them in the language they can best understand. In fact, the Improving America's Schools Act has mandated that "LEP students shall be assessed relative to State content standards, to the extent possible, in the language and form most likely to yield accurate/reliable information on what students know and can do, to determine mastery of skills in subjects other than English" (sec. 1111(b)(3)(F)(i-iii)). It is important to remember that LEP students may not necessarily perform better when tested in their native language if most of their schooling has been in English. Information on students' relative proficiency in English and the native language and on prior school experience should help determine the best language of assessment.

Because Spanish speakers comprise by far the largest LEP student population in the country, several instruments have been published in that language. Some of those are described here. This discussion is not intended to recommend these instruments, but simply to make the reader aware of them. Technical or examiner manuals are available from the publisher with information on content coverage and grade levels tested, and detailed instructions on administration, scoring, and interpretation. Readers who are interested in a specific instrument should request additional information from the publisher and determine for themselves whether it suits their programs.

Aprenda: La Prueba de Logros en Español
Published by The Psychological Corporation (Harcourt Brace)

Aprenda has several forms, covering pre-primer, primary, and intermediate grades. It generally matches the objectives of the Stanford Achievement Tests. Different forms provide different coverage according to their grade levels, but in general, Aprenda measures reading, language arts, and mathematics.

Integrated Assessment System (IAS)
Published by The Psychological Corporation (Harcourt Brace)

The IAS is a set of performance-based tests in language arts and a language arts portfolio for grades K-8. The tasks involve both reading and writing and are intended to elicit higher-order thinking.

La Prueba Riverside de Realización en Español
Published by Riverside Publishing (Houghton Mifflin)

La Prueba is available in different levels, which should be selected according to students' Spanish literacy. Topics assessed include reading comprehension, vocabulary, study skills, grammar, math computation and problem solving, social studies, and science.

Riverside Performance Assessment Series (RPAS)
Published by Riverside Publishing (Houghton Mifflin)

This series contains performance assessments for mathematics, reading, and writing. They are available in parallel English and Spanish forms. The Spanish edition has four levels, for grades three through five, four through six, six through nine, and nine through 12.

Batería Woodcock-Muñoz: Pruebas de aprovechamiento—Revisada
Published by Riverside Publishing (Houghton Mifflin)

The Batería Woodcock-Muñoz: Pruebas de aprovechamiento is the Spanish parallel version of the Woodcock-Johnson-Revised Tests of Achievement. It spans all grade levels, kindergarten through college-educated. It gives information in clusters of reading, mathematics, written language, and general knowledge. It is individually administered.

Woodcock Language Proficiency Battery—Revised
Published by Riverside Publishing (Houghton Mifflin)

This instrument assesses language skills in Spanish; it also exists in English. The Woodcock-Muñoz Language Survey, described in the discussion of language proficiency assessment, is drawn from this larger battery, which assesses 16 different components of oral language, reading, and written language. It spans all grade levels. It is individually administered.

Spanish Assessment of Basic Education (SABE)
Published by CTB\MacMillan\McGraw-Hill

The SABE is a series of norm-referenced tests for grades one through eight designed to measure achievement in the basic skills of reading, mathematics,

spelling, language, and study skills. It is statistically linked with the Comprehensive Tests of Basic Skills (CTBS) and the California Achievement Tests (CAT).

Test de Vocabularío en Imagenes Peabody: Adaptación Hispanomerica
Published by American Guidance Service

This is a Spanish version of the Peabody Picture Vocabulary Test. It spans ages two through 18. It can be used as an achievement test to show progress in Spanish. It has also been used as an indicator of language proficiency or a screening test of scholastic aptitude. It is individually and orally administered. Spanish norms have not yet been developed.

Special Education

Assessment for special education poses a delicate dilemma for many districts. Legally and ethically, students should not be referred for special education services simply because language differences prevent them from learning effectively. However, LEP students who have a disability must be so identified and served. Special education personnel as well as bilingual and ESL program personnel should work together to devise pre-referral and identification strategies that combine the unique insights of both. Some of the achievement instruments listed previously are frequently used in the screening process. Several other instruments commonly used in special education identification have been produced in Spanish versions and are listed below. Also listed are instruments and procedures that can be used in the pre-referral process.

The Acculturation Quick Screen (AQS)
CrossCultural Developmental Education Services

The AQS is a means of measuring the relative level of acculturation to public school culture in the United States. It is based on research on the factors that predict the degree of successful integration for those who are experiencing culture shock. It is intended as part of the initial information-gathering process for making instructional decisions about culturally and linguistically diverse students and as part of the intervention planning stages that should occur prior to making a formal referral for special services. The AQS is available in either a protocol, a tablet of scoring forms, or a software version, the Acculturation Quick Screen Wizard (AQSW).

Batería Woodcock-Muñoz: Pruebas de habilidad cognitiva—Revisada
Published by Riverside Publishing (Houghton Mifflin)

The Batería Woodcock-Muñoz is the Spanish version of the Woodcock-Johnson-Revised Tests of Cognitive Ability. The Spanish and English versions are parallel and therefore can provide comparative language information. The tests are designed to assess long-term retrieval, short-term memory, processing speed, auditory processing, visual processing, comprehension-knowledge, and fluid reasoning.

Boehm Test of Basic Concept—Revised, Spanish Edition
Published by The Psychological Corporation (Harcourt Brace)

This is the Spanish edition of the Boehm Test, designed to measure mastery of concepts fundamental to understanding verbal instruction and necessary for early school achievement. It helps identify students with basic concept deficiencies. It can be administered individually or to small groups. Spanish norms are not available for this test.

Escala de Inteligencia Wechsler para Niños
Published by The Psychological Corporation (Harcourt Brace)

This is the Spanish translation of the Wechsler Intelligence Scale for Children, intended for use with ages five through 15.

The Sociocultural Checklist
CrossCultural Developmental Education Services

The Sociocultural Checklist was developed as an initial screening tool for educators in American public schools who are concerned about the learning and behavior of a specific student from a culturally or linguistically diverse background. The Sociocultural Checklist is scored by a teacher or team of education personnel who are familiar with the student's background and classroom behavior. It is recommended as part of the initial information-gathering and intervention-planning stages that should occur prior to making a formal referral for special services.

Conclusion

Setting up and executing an assessment program for an ESL or bilingual program does not follow a simple formula. As this paper shows, it takes consideration of many factors, and a comprehensive assessment program involves staff at many different levels. The task may seem daunting at first, but like anything else, it

becomes easier with practice and familiarity. LEP students are an increasing presence in American schools in every region of the country, and concomitantly, there is increasing interest in second-language acquisition and effective alternative programs. At the same time, accountability has taken on new importance, and educators who may have gritted their teeth through their college tests and measurements classes have become excited about high standards, portfolios, and performance assessment. Combined interest and expertise in the education of language-minority students and the assessment of learning behoove any ESL or bilingual program.

Persons interested in the content of this paper should pursue that interest. The attached bibliography lists books, articles, and monographs dealing with the assessment of limited-English-proficient students. No entries are included for testing theory or data analysis because those topics are generic to education. A significant amount of information can be found online through the Internet, including more detailed information on the published tests listed in this paper. Useful key words include NCBE, TESOL, and ERIC/AE Test Locator.

Bibliography

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge, England: Cambridge University Press.

Alderson, J. C., Krahnke, K. J., & Stansfield, C. W. (1987). *Review of English language proficiency tests.* Washington, DC: TESOL.

Collier, C. (1988). *Assessing minority students with learning and behavior problems.* Boulder, CO: Hamilton Publications.

Collier, C. (n.d.). *Separating difference from disability: Assessing diverse learners.* Boulder, CO: Hamilton Publications.

Council of Chief State School Officers. (1992). *Recommendations for improving the assessment and monitoring of students with limited English proficiency.* Alexandria, VA: Council of Chief State School Officers.

Cumming, A., & Berwick, R. (Eds.). (1996). *Validation in language testing.* Clevedon, England: Multilingual Matters.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy.* Clevedon, England: Multilingual Matters.

Del Vecchio, A., & Guerrero, M. (1995). *Handbook of English language proficiency tests.* Washington, DC: National Clearinghouse on Bilingual Education.

Dolson, D. (1994). *Assessing students in bilingual contexts: Provisional guidelines.* Sacramento, CA: California Department of Education. (ERIC Document Reproduction Service No. ED 373 581)

Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education.* Cambridge, England: Cambridge University Press.

Hafner, A. L. (1994). *Making our assessments comprehensible to English learners.* Paper presented at the annual CCSSO Large-Scale Assessment Conference, Albuquerque, NM.

Harley, B., Allen, P., Cummins, J., & Swain, M. (Eds.). (1990). *The development of second language proficiency.* Cambridge, England: Cambridge University Press.

Hoover, J. J., & Collier, C. (1988). *Educating minority students with learning and behavior problems: Strategies for assessment and teaching.* Boulder, CO: Hamilton Publications.

Hughes, A. (1989). *Testing for language teachers.* Cambridge, England: Cambridge University Press.

Navarrete, C., Wilde, J., Nelson, C., Martinez, R., & Hargett, G. (1990, summer). *Informal assessment in educational evaluation: Implications for bilingual education programs.* Washington, DC: NCBE. (ERIC Document Reproduction Service No. ED 337 041)

O'Malley, J. M., & Pierce, L. V. (1991, November). *Using portfolio and alternative assessment with LEP students: Program information guide* (No. 9). Washington, DC: NCBE. (ERIC Document Reproduction Service No. ED 346 747)

O'Malley, J. M., & Pierce, L. V. (1997). *Authentic assessment for English language learners: Practical approaches for teachers.* Reading, MA: Addison Wesley Longman.

Palmer, A. S., Groot, P. J. M., & Trosper, G. A. (Eds.). (1981). *The construct validation of tests of communicative competence.* Washington, DC: TESOL.

Roos, P. (1995). *Rights of limited English proficient students under federal law—A guide for school administrators.* Unpublished paper presented at Weber State University, Success for all Students Conference, Ogden, UT.

Valdes, G., & Figueroa, R. (1994). *Bilingualism and testing: A special case of bias.* Norwood, NJ: Ablex Publishing.

Northwest Regional Educational Laboratory's
Comprehensive Center, Region X
101 S.W. Main St., Suite 500
Portland, OR 97204